TEMOS: Generating diverse human motions from textual descriptions

Year: 2022 | Citations: 499 | Authors: Mathis Petrovich, Michael J. Black, Gul Varol

---

Abstract

We address the problem of generating diverse 3D human motions from textual descriptions. This challenging task requires joint modeling of both modalities: understanding and extracting useful human-centric information from the text, and then generating plausible and realistic sequences of human poses. In contrast to most previous work which focuses on generating a single, deterministic, motion from a textual description, we design a variational approach that can produce multiple diverse human motions. We propose TEMOS, a text-conditioned generative model leveraging variational autoencoder (VAE) training with human motion data, in combination with a text encoder that produces distribution parameters compatible with the VAE latent space. We show the TEMOS framework can produce both skeleton-based animations as in prior work, as well more expressive SMPL body motions. We evaluate our approach on the KIT Motion-Language benchmark and, despite being relatively straightforward, demonstrate significant improvements over the state of the art. Code and models are available on our webpage.